



---

Year: 2020

---

## The Genetic Landscape and Epidemiology of Phenylketonuria

Hillert, Alicia ; Anikster, Yair ; Belanger-Quintana, Amaya ; Burlina, Alberto ; et al ; Rohrbach, Marianne ; Thöny, Beat ; Blau, Nenad

**Abstract:** Phenylketonuria (PKU), caused by variants in the phenylalanine hydroxylase (PAH) gene, is the most common autosomal-recessive Mendelian phenotype of amino acid metabolism. We estimated that globally 0.45 million individuals have PKU, with global prevalence 1:23,930 live births (range 1:4,500 [Italy]–1:125,000 [Japan]). Comparing genotypes and metabolic phenotypes from 16,092 affected subjects revealed differences in disease severity in 51 countries from 17 world regions, with the global phenotype distribution of 62% classic PKU, 22% mild PKU, and 16% mild hyperphenylalaninemia. A gradient in genotype and phenotype distribution exists across Europe, from classic PKU in the east to mild PKU in the southwest and mild hyperphenylalaninemia in the south. The c.1241A>G (p.Tyr414Cys)-associated genotype can be traced from Northern to Western Europe, from Sweden via Norway, to Denmark, to the Netherlands. The frequency of classic PKU increases from Europe (56%) via Middle East (71%) to Australia (80%). Of 758 PAH variants, c.1222C>T (p.Arg408Trp) (22.2%), c.1066–11G>A (IVS10–11G>A) (6.4%), and c.782G>A (p.Arg261Gln) (5.5%) were most common and responsible for two prevalent genotypes: p.[Arg408Trp];[Arg408Trp] (11.4%) and c.[1066–11G>A];[1066–11G>A] (2.6%). Most genotypes (73%) were compound heterozygous, 27% were homozygous, and 55% of 3,659 different genotypes occurred in only a single individual. PAH variants were scored using an allelic phenotype value and correlated with pre-treatment blood phenylalanine concentrations (n = 6,115) and tetrahydrobiopterin loading test results (n = 4,381), enabling prediction of both a genotype-based phenotype (88%) and tetrahydrobiopterin responsiveness (83%). This study shows that large genotype databases enable accurate phenotype prediction, allowing appropriate targeting of therapies to optimize clinical outcome.

DOI: <https://doi.org/10.1016/j.ajhg.2020.06.006>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-194383>

Journal Article

Accepted Version



The following work is licensed under a Creative Commons: Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.

Originally published at:

Hillert, Alicia; Anikster, Yair; Belanger-Quintana, Amaya; Burlina, Alberto; et al; Rohrbach, Marianne; Thöny, Beat; Blau, Nenad (2020). The Genetic Landscape and Epidemiology of Phenylketonuria. *American Journal of Human Genetics*, 107(2):234-250.

DOI: <https://doi.org/10.1016/j.ajhg.2020.06.006>

# The genetic landscape and epidemiology of phenylketonuria

Alicia Hillert<sup>1</sup>, Yair Anikster<sup>2</sup>, Amaya Belanger-Quintana<sup>3</sup>, Alberto Burlina<sup>4</sup>, Barbara K. Burton<sup>5</sup>, Carla Carducci<sup>6</sup>, Ana E Chiesa<sup>7</sup>, John Christodoulou<sup>8</sup>, Maja Đorđević<sup>9</sup>, Lourdes R. Desviat<sup>10</sup>, Aviva Eliyahu<sup>2</sup>, Roeland A.F. Evers<sup>11</sup>, Lena Fajkusova<sup>12</sup>, François Feillet<sup>13</sup>, Pedro Eduardo Bonfim Freitas<sup>14</sup>, Maria Giżewska<sup>15</sup>, Polina Gundorova<sup>16</sup>, Daniela Karall<sup>17</sup>, Katya Kneller<sup>2</sup>, Sergey I. Kutsev<sup>16</sup>, Vincenzo Leuzzi<sup>18</sup>, Harvey L. Levy<sup>19</sup>, Uta Lichter-Konecki<sup>20</sup>, Ania C. Muntau<sup>21</sup>, Fares Namour<sup>13</sup>, Mariusz Oltarzewski<sup>22</sup>, Andrea Paras<sup>5</sup>, Belen Perez<sup>10</sup>, Emil Polak<sup>23</sup>, Alexander V. Polyakov<sup>16</sup>, Francesco Porta<sup>24</sup>, Marianne Rohrbach<sup>25</sup>, Sabine Scholl-Bürgi<sup>17</sup>, Norma Spécola<sup>26</sup>, Maja Stojiljković<sup>27</sup>, Nan Shen<sup>28</sup>, Luiz Carlos Santana da Silva<sup>14</sup>, Anastasia Skouma<sup>29</sup>, Francjan van Spronsen<sup>11</sup>, Vera Stoppioni<sup>30</sup>, Beat Thöny<sup>25</sup>, Friedrich K. Trefz<sup>1</sup>, Jerry Vockley<sup>20</sup>, Youngguo Yu<sup>31</sup>, Johannes Zschocke<sup>32</sup>, Georg F. Hoffmann<sup>1</sup>, Sven Garbade<sup>1\*\*</sup>, Nenad Blau<sup>1,25\*</sup>

<sup>1</sup>Dietmar-Hopp Metabolic Center, University Children's Hospital, 69120 Heidelberg, Germany; <sup>2</sup>Metabolic Disease Unit, Edmond and Lily Safra Children's Hospital, Sheba Medical Center, Tel-Aviv University, 52621 Tel-Aviv, Israel; <sup>3</sup>Unidad de Enfermedades Metabólicas, Servicio de Pediatría, Hospital Ramon y Cajal, 28034 Madrid, Spain; <sup>4</sup>Division of Inherited Metabolic Diseases, Department of Woman's and Child's Health, University Hospital, 35129 Padua, Italy; <sup>5</sup>Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, IL 60611, USA; <sup>6</sup>Department of Experimental Medicine, Sapienza University of Rome, 00185 Rome, Italy; <sup>7</sup>Fundación de Endocrinología Infantil (FEI), C1425 Buenos Aires, Argentina; <sup>8</sup>Murdoch Children's Research Institute and Department of Pediatrics, University of Melbourne, Melbourne, Victoria 3052, Australia; <sup>9</sup>Institute of Mother and Child Healthcare "Dr. Vukan Čupić", 11000 Belgrade, Serbia; <sup>10</sup>Centro de Diagnóstico de Enfermedades Moleculares, Centro de Biología Molecular CSIC-UAM, Universidad Autónoma de Madrid. CIBERER, IdiPAZ, 28049 Madrid, Spain; <sup>11</sup>University of Groningen, University Medical Center Groningen, Beatrix Children's Hospital, Section of Metabolic Diseases, 9712 CP Groningen, the Netherlands; <sup>12</sup>Centre of Molecular Biology and Gene Therapy, University Hospital Brno, 62500 Brno, Czech Republic; <sup>13</sup>Reference Center for Inherited Metabolic Diseases, University Hospital of Nancy, 54511 Vandoeuvre-lès-Nancy, France; <sup>14</sup>Laboratory of Inborn Errors of Metabolism, Institute of Biological Sciences, Federal University of Pará, Belém 66075-110, Brazil; <sup>15</sup>Department of Pediatrics, Endocrinology, Diabetology, Metabolic Diseases and Cardiology, Pomeranian Medical University, 71-252 Szczecin, Poland; <sup>16</sup>Research Centre for Medical Genetics, 115522 Moscow, Russia; <sup>17</sup>Clinic of Pediatrics,

Division of Inherited Metabolic Disorders, Medical University of Innsbruck, 6020 Innsbruck, Austria; <sup>18</sup>Department of Human Neuroscience, Sapienza University of Rome, 00185 Rome, Italy; <sup>19</sup>Division of Genetics and Genomics, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA; <sup>20</sup>UPMC, Children's Hospital of Pittsburgh, Pittsburgh, PA 15224, USA; <sup>21</sup>University Children's Hospital, University Medical Center Hamburg Eppendorf, 20246 Hamburg, Germany; <sup>22</sup>Department of Screening and Metabolic Diagnostics, Institute of Mother and Child, 01-211 Warsaw, Poland; <sup>23</sup>Comenius University, Faculty of Natural Sciences, Department of Molecular Biology, 84215 Bratislava 4, Slovak Republic; <sup>24</sup>Department of Pediatrics, AOU Citta' della Salute e della Scienza di Torino, 10126 Torino, Italy; <sup>25</sup>Division of Metabolism, University Children's Hospital, 8032 Zürich, Switzerland; <sup>26</sup>Unidad de Metabolismo. Hospital de Niños "Sor Ludovica" de La Plata, 1904 Buenos Aires, Argentina; <sup>27</sup>Institute of Molecular Genetics and Genetic Engineering, University of Belgrade, 11000 Belgrade, Serbia; <sup>28</sup> Department of Infectious Diseases, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, 2000025 Shanghai, China; <sup>29</sup>Institute of Child Health, 11526 Athens, Greece; <sup>30</sup>Centro Screening Neonatale Regione Marche, Azienda Ospedaliera Ospedali Riuniti Marche Nord, 61032 Fano, Italy; <sup>31</sup>Department of Pediatric Endocrinology/Genetics, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai Institute for Pediatric Research, 2000025 Shanghai, China; <sup>32</sup>Institute of Human Genetics, Medical University Innsbruck, 6020 Innsbruck, Austria.

\*Correspondence:

Nenad Blau, PhD  
 Professor of Biochemical Genetics  
 Dietmar-Hopp Metabolic Center  
 University Children's Hospital  
 69120 Heidelberg  
 Germany  
[nenad.blau@med.uni-heidelberg.de](mailto:nenad.blau@med.uni-heidelberg.de)  
[nenad.blau@kispi.uzh.ch](mailto:nenad.blau@kispi.uzh.ch)

\*\*Co-correspondence: [sven.garbade@med.uni-heidelberg.de](mailto:sven.garbade@med.uni-heidelberg.de)

## Abstract

Phenylketonuria (PKU), caused by variants in the phenylalanine hydroxylase (PAH) gene, is the most common autosomal recessive Mendelian phenotype of amino acid metabolism. We estimated that globally 0.45 million individuals have PKU, with global prevalence 1:23,930 live births (range 1:4,500 [Italy] - 1:125,000 [Japan]). Comparing genotypes and metabolic phenotypes from 16,092 patients revealed differences in diseases severity in 51 countries from 17 world regions, with the global phenotype distribution of 62% classic PKU, 22% mild PKU and 16% mild hyperphenylalaninemia. A gradient in genotype and phenotype distribution exists across Europe, from classic PKU in the east, to mild PKU in the southwest and mild hyperphenylalaninemia in the south. The c.1241A>G (p.Tyr414Cys)-associated genotype can be traced from Northern to Western Europe, from Sweden via Norway, Denmark to the Netherlands. The frequency of classic PKU increases from Europe (56%) via Middle East (71%) to Australia (80%). Of 758 *PAH* variants, c.1222C>T (p.Arg408Trp) (22.2%), c.1066-11G>A (IVS10-11G>A) (6.4%) and c.782G>A (p.Arg261Gln) (5.5%) were most common and responsible for two prevalent genotypes: p.[Arg408Trp];[Arg408Trp] (11.4%) and c.[1066-11G>A];[1066-11G>A] (2.6%). Most genotypes (73%) were compound heterozygous, 27% were homozygous and 55% of 3,659 different genotypes occurred in only a single patient. *PAH* variants were scored using an allelic phenotype value and correlated with pre-treatment blood phenylalanine concentrations (n=6,115) and tetrahydrobiopterin loading test results (n=4,381), enabling prediction of both a genotype-based phenotype (88%) and tetrahydrobiopterin responsiveness (83%). This study shows that large genotype databases enable accurate phenotype prediction, allowing appropriate targeting of therapies to optimize clinical outcome.

## Introduction

Phenylketonuria (PKU; MIM 261600) is the most frequent inborn error of the amino acid metabolism. More than 1,180 biallelic variants in the phenylalanine hydroxylase (*PAH*) gene located on chromosome 12q22-24.1 have been identified.<sup>1</sup> These autosomal recessive inherited variants lead to deficiency in the *PAH* enzyme, which hydroxylates phenylalanine to tyrosine, with the help of a cofactor (tetrahydrobiopterin; BH<sub>4</sub>), molecular oxygen, and non-heme iron.<sup>2</sup>

The metabolic picture is highly heterogeneous as it depends on the degree of residual *PAH* activity and blood phenylalanine (Phe) concentrations. Lower residual enzyme activity usually results in higher blood Phe concentrations and a more severe clinical phenotype if left untreated.<sup>3</sup>

The severity of PKU is defined by daily Phe tolerance.<sup>4</sup> Standard classification is according to the pre-treatment blood Phe concentration and daily dietary Phe tolerance, ranging from the severe classical PKU (cPKU) with pre-treatment blood Phe concentrations of >1,200 µmol/L, to mild PKU (mPKU) with pre-treatment blood Phe concentrations of 600–1,200 µmol/L and mild hyperphenylalaninemia (MHP) with pre-treatment Phe blood concentrations of 120–600 µmol/L.<sup>1; 5</sup>

Untreated PKU generally results in global developmental delay or severe irreversible intellectual disability, as well as growth failure, hypopigmentation, motor deficits, ataxia, and seizures.<sup>1</sup> The population of PKU patients is heterogeneous in terms of treatment history and diet compliance.<sup>5; 6</sup> Early diagnosis and treatment with a low-Phe diet has enabled an almost normal life for the majority of PKU patients.<sup>7</sup> Pharmacological treatment with BH<sub>4</sub> (sapropterin) and enzyme substitution therapy with Phe ammonia lyase (PAL) provide alternative treatment options for some patients.<sup>8</sup>

PKU is one of the most frequent inherited disorders in Caucasians, with an incidence of roughly 1:10,000 live births in the USA,<sup>9</sup> although the prevalence of PKU varies significantly among ethnicities and geographic regions worldwide. In Europe, the incidence of PKU ranges from 1: 850 in the Karachay-Cherkess Republic (Russia)<sup>10</sup> to only 1:112,000 live births in Finland.<sup>11</sup> PKU occurs less often in Japan, with an incidence of 1:125,000.<sup>12</sup>

A large number of *PAH* variants give rise to a wide scale of residual *PAH* enzyme activities that correspond to different PKU phenotypes.<sup>13; 14</sup> Associations between genotypes and *in vitro* residual *PAH* activity have been documented for many *PAH* variants.<sup>3</sup> Therefore, the molecular genetics of PKU and genotype-based phenotype prediction may be clinically

useful, particularly where treatment recommendations are unclear (e.g. due to borderline blood phenylalanine concentrations) or for genetic counseling of patients' families.

The major goal of this study was the analysis of a large database of PKU phenotype and genotype to elucidate the current distribution of PKU worldwide and to create an overview of the severity of *PAH* variants, genotypes and the resulting phenotypes, in various geographic regions and respective countries. Furthermore, this work improves the accuracy of genotypic phenotype prediction by the use of the allelic phenotype value (APV) value and the genotypic BH<sub>4</sub> responsiveness prediction.

## Subjects and methods

### Literature search

An electronic search using the databases MEDLINE (via Pub Med), the Cochrane library and Web of Science was carried out to compare articles that were published between January 1980 and October 2019, covering the epidemiology and genetics of PKU in different world regions. Key words included Phenylketonuria[MeSH Terms] OR "phenylalanine hydroxylase/deficiency"[MeSH Terms] OR hyperphenylalaninaemia[Title/Abstract] OR hyperphenylalaninemia[Title/Abstract] OR PAH deficiency[Title/Abstract] OR phenylalanine hydroxylase deficiency[Title/Abstract] OR phenylketonuri\*[Title/Abstract] OR pku[Title/Abstract]) AND ( "1980/01/01"[PDat] : "2019/12/31"[PDat] ) AND Humans[Mesh] Filters: Publication date from 1980/01/01 to 2019/12/31; Humans.

This literature search yielded 5,459 records without duplicates, of which 1,118 papers with an appropriate title and abstract were assessed. The final number of relevant records was 256. (Figure S1)

### Databases

The *PAH* locus-specific database *PAH*vdb, ClinVar, HGMD and LOVD databases were searched for variants. *PAH*vdb is linked to the genotype-phenotype BIOPKU database and was used for analyses. The RefSeq accession number for the *PAH* is ENSG00000171759; GeneBank: NM\_000277.1. All variants were tested using Mutalyzer 2.0 and follow the HGVS guidelines.

The BIOPKU database encompasses information about more than 16,900 PKU patients from 51 countries, providing information on patient genotypes, corresponding metabolic phenotypes, BH<sub>4</sub> responsiveness (where reported) and highest blood Phe concentrations

before starting treatment (where reported). Individual information was collected from the published literature or anonymized records submitted online. Phenotype information was unknown for 690 patients. Data in the database are anonymized and cannot be traced back to the families. Purpose of the database was to provide an online search tool for reported *PAH* variants or genotypes, with an output summarizing the number of records, phenotype distribution, BH<sub>4</sub> responsiveness and regional and counties distribution. As mentioned above, BIOPKU is linked with the *PAH*vdb, thus providing additional information incl. allelic phenotype values (APV). Data submissions and procedures followed were in accordance with the ethical standards and were approved by local institutional review boards where applicable. Table 1 shows the information included in the database.

### **Definition of phenotypes**

Since not all countries use the same nomenclature for the severity of PAH deficiency, in this study the following 3 metabolic phenotype groups were used: classical PKU (cPKU; pre-treatment blood Phe >1,200 µmol/L); mild PKU (mPKU; pre-treatment blood Phe 600–1,200 µmol/L) and mild hyperphenylalaninemia (MHP; pre-treatment blood Phe 120–600). Patients with a moderate PKU (pre-treatment blood Phe 900–1,200 µmol/L) were included in the mPKU group and MHP included MHP-no treatment (blood Phe 120–360 µmol/L) and MHP-gray zone (360–600 µmol/L) categories.<sup>5-7</sup> Any classifications, not fitting into one of the above groups (due to different country-specific classifications), were reassigned on the basis of reported pre-treatment blood Phe concentrations. However, neither the method used for blood Phe quantification nor the age at measurement were reported.

### **Definition of the allelic (APV) and genotypic phenotype value (GPV)**

APV is a value defining the association of a variant with the corresponding metabolic phenotype, thus defining its severity. APV was calculated for variants occurring in a functionally hemizygous constellation (i.e. in a combination with an inactive null allele) in at least 5 patients.<sup>15</sup> APVs range between 0 and 10, with following classifying definitions cPKU (APV = 0-2.7), mPKU (APV = 2.8-6.6) and MHP (APVs 6.7-10).<sup>15</sup>

The genotypic phenotype value (GPV) was calculated from the APVs of both alleles and was assigned to a higher APV (APV<sub>max</sub>). This calculation was based on the fact that the milder variant (with a higher APV) is always dominant over the severe one.<sup>15; 16</sup> Possible effects of interallelic complementation and epigenetic factors, which may influence the phenotype<sup>17; 18</sup>, were not considered in this study.



## **Definition of BH<sub>4</sub> responsiveness**

BH<sub>4</sub> responsiveness was defined as a  $\geq 30\%$  reduction of blood Phe concentrations within 24-48 hours after the administration of BH<sub>4</sub> (20mg/kg body weight)<sup>19; 20</sup>. The protocols for the BH<sub>4</sub> challenge were different in different centers (i.e. 24h, 48h, 1 day, 1 week, etc.) and for this reason BH<sub>4</sub> responsiveness was simply reported as “yes” or “no”. A linear discriminant analysis was applied to predict BH<sub>4</sub> responsiveness based on the APV for untested patients.

## **Statistical analysis**

Statistical analysis was performed using R, an open source software and flexible programming language used for the statistical data analysis as well as graphic creations (<https://www.r-project.org/>). A total of 16,196 records with a complete genotype information (variant 1 and 2 known) were analyzed. Two linear discriminant analyses (LDA) were computed: first LDA was computed to predict the clinical phenotype (cPKU, mPKU and MHPA) from individual's GPV, the second LDA was computed to predicted BH<sub>4</sub> responsiveness by GPV value from untested individuals in BIOPKU.

## **Results**

### **Prevalence of PKU**

Based on the literature search and reports from national screening centers, the prevalence of PKU was estimated for 64 countries. For parts of Africa, Asia, South America and Caribbean there was no information. The estimated total number of PKU patients (all phenotypes) from those 64 countries in 2018 was 360,466. For the remaining 257 countries we were unable to find credible PKU prevalence sources; we used the average regional prevalence of 64 countries, multiplied by their populations, resulting in an additional 94,114 PKU patients (total about 0.45 million PKU patients). The global PKU prevalence was estimated to be 1:23,930 newborns (Figure 1, Table S1).

The PKU prevalence was highest in European and certain Middle Eastern populations. Italy (1:4,000) and Ireland (1:4,545) had even higher prevalence than Iran and Jordan (both 1:5,000) or Turkey (1:6,667). However, Saudi Arabia (1:14,245), Iraq (1:14,286) or United Arab Emirates (1:14,493) had lower PKU prevalence.

PKU prevalence was also high in Central European countries, e.g. Germany (1:5,360), Czechia (1:5,521), Austria (1:5,764) and Slovakia (1:5,753). Slovenia (1:7,143) and Poland

(1:8,309) had similar rates to Eastern Europe, e.g. Estonia (1:7,143), Russia (1: 7,714) Belarus (1:7,692), and Croatia (1:8,333).

PKU occurred slightly less frequently in Western Europe, e.g. in France (1:9,091), United Kingdom (1:10,000), Belgium (1:11,000) or the Netherlands (11,546), and in Southern Europe, e.g. Spain (1:10,115) or Portugal (1:12,500).

Northern Europe showed the lowest PKU rates in Europe, e.g. Norway (1:11,457), Sweden (1: 12,681), Denmark (1: 13,434), or only 1:112,000 in Finland.

PKU occurred more frequently In Canada (1:15,000) than in the United States of America (1:25,000) or in Latin American countries, e.g., Argentina (1:15,715), Chile (1:19,231), Brazil (1:25,000), Mexico (1:27,778) or Peru (1:46,970).

The lowest PKU prevalence was reported in Asian countries, such as Thailand (1:227,273), Japan (1:125,000), Philippines (1:116,006) or Singapore (1:83,333). One exception was China where the PKU prevalence was 1:15,924, which was comparable to Europe.

### **Descriptive analysis of the phenylalanine hydroxylase gene locus-specific database**

Substitutions were by far the most frequent variant type in the *PAH*vdB (80.5%), followed by deletions (12.9%) and duplications (2.1%). Of all variants, 691 were missense variants (58.3%), followed by 165 frameshift variants (13.9%) and 155 splice site (13.1%) variants. Nonsense variants (6.9%), synonymous variants (4.9%) and in-frame variants (1.9%) were less frequent. Extension, complex and unknown variants, accounted for the remainder. Exon 6 contained the largest number of variants (14.1%), followed by exon 7 (12.2%) and exon 3 (9.9%). Most variants (59.2%) were located in the central catalytic domain, 17.5% in the N-terminal regulatory domain and 5.4% in the C-terminal oligomerization domain of the PAH monomer. The remaining variants (17.9%) were either in the intronic or UTR regions. Only 7.7% of all variants were located in one of the 4 cofactor binding regions (Figure S2).

The APV was known for 589 of 1,186 variants. Most variants (441) were defined as severe null alleles (APV=0), 32 variants as cPKU phenotype, 52 as mPKU and 64 as MHP alleles. Metabolic phenotype was predicted, on the basis of the known genotype and corresponding GPV from the LDA, in 87.9% of PKU patients.

### **Descriptive analysis of the BIOPKU database**

As of October 2019, the BIOPKU database contained anonymized data on more than 16,900 PKU patients. A total of 16,196 PKU patients with 3,659 different genotypes were analyzed. Information on the PKU phenotype was available for 16,902 patients, thus some of them

(810) with no genotype information. Maximum pretreatment Phe concentrations were reported for only 6,371 patients and the BH<sub>4</sub> test information was available for 5,597 patients (Table 1).

### **Global PKU phenotype and genotype distribution**

Information on the PKU phenotype and genotype was available for 16,092 out of 16,196 patients (99%). Of these, most had cPKU (9,923; 61.7%), 3,521 (21.9%) had mPKU, and 2,648 had MHP (16.4%). Information about the phenotype was unavailable for 104 patients. The comparison of pre-treatment Phe level with the reported phenotype (n=6,369) is illustrated in Figure 2A and with the GPVs (n=6,115) in Figure 2B. The interquartile range (n, median, 25<sup>th</sup>–75<sup>th</sup> percentile) was smallest and lowest for the MHP patient group: 1,283, 320 µmol/L, 242–432 µmol/L, and was larger for mPKU and cPKU (1,487, 793 µmol/L, 660–793 µmol/L, and 3,599, 1,550 µmol/L, 1,270–1936 µmol/L, respectively).

A total of 758 different *PAH* variants were identified in this study. The three most prevalent variants were c.1222C>T (p.Arg408Trp), with an allele frequency (AF) = 22.2%, c.1066-11G>A (IVS10-11G>A; AF=6.4%) and c.782G>A (p.Arg261Gln; AF 5.5%). Figure 4 shows the most frequent phenotype-specific variants. A full list of *PAH* variants is shown in Table S2.

Of all patients, 11,810 (72.9%) were compound heterozygotes and 4,386 (27.1%) were homozygotes. Of 3,659 genotypes, 3,446 (94.2%) were compound heterozygotes and 213 (5.8%) homozygotes. The three most prevalent genotypes were: p.[Arg408Trp];[Arg408Trp] with a genotype frequency (GF) =11.4%, followed by c.[1066-11G>A];[1066-11G>A] (GF=2.6%) and c.[1222C>T];[1315+1G>A] (p.[Arg408Trp];IVS12+1G>A) (GF=1.6%). A full list of genotypes is shown in Table S3. Strikingly, 54.5% of all genotypes were specific for only one patient and were not used for the phenotype prediction.

### **BH<sub>4</sub> responsiveness**

Information on BH<sub>4</sub> responsiveness was available for 5,597 patients. About half were classified as BH<sub>4</sub> responsive (2,316; 43%). Table S4 shows the distribution of BH<sub>4</sub> responsiveness in different phenotype groups. As expected, milder forms of PKU appeared more likely to be BH<sub>4</sub> responsive, whereas most patients with cPKU were non-responders. Pretreatment blood Phe concentrations (n, median, 25<sup>th</sup>–75<sup>th</sup> percentile) were much lower for BH<sub>4</sub> responsive patients (1,116, 620 µmol/L, 411–853 µmol/L), compared with non-responders (1,180, 1,361 µmol/L, 1,039–1,719 µmol/L (Figure 2C). The 11,584 patients not

tested for BH<sub>4</sub> responsiveness were analyzed for potential responsiveness using the GPVs. Patients with an GPV>3.8 (n=3,023; 26%) were assigned as potential BH<sub>4</sub> responders ( $p<0.001$  vs. patients assigned as non-responders [n=8,561; 74%]) (Figure 2D).

## **PKU phenotypes and genotypes in world regions**

### ***Overview of global data***

Patient data originated from 51 countries from 17 world regions and a total of 15,357 patients from 33 countries with at least 35 reported cases were analyzed (Tables S5 and S6, Figure S3). Most cases were reported from Central (18.3%) and Eastern (18.1%) Europe, Eastern Asia (13.8%), Western Europe (9.6%) and the Middle East (9.2%). Very few patients were reported from the northern and southern parts of Africa.

cPKU was the most frequent phenotype in all world regions, with high rates reported in Australia and eastern Europe (81%), and rates of <50% reported only in Serbia, Argentina, Turkey, Netherlands, Sweden, Spain, Italy, Japan, Slovenia, Germany, and Taiwan (Figure 4A). Of 2,835 patients in Russia, 81% had cPKU, 9.9 % had mPKU and 9.1% had MHP. Estonia, representing the Baltic region, reported 93.5% of patients with cPKU. The ratio of mPKU:MHP patients was comparable, except for Eastern Asia, North America and South America, where mPKU was more common. The phenotype distribution in world regions, as well as Genotypic Phenotype Values (GPV) for Europe and other world regions is shown in a more granular form in Figure S4.

Reports from individual regions are summarized in Table 2 and below.

### ***Europe***

Variant p.Arg408Trp was the most common (AF=63.7%) throughout Europe, followed by c.1066-11G>A (AF=11%) and p.Arg261Gln (AF=11%). Eastern Europe had the highest AF for p.Arg408Trp (54.6%), mostly with homozygous genotype p.[Arg408Trp];[Arg408Trp] (GF=32.7%); findings were similar for Central Europe (AF=44.4% and GF=23.8%). Russia contributed the largest number of records. A single variant, p.Arg408Trp (AF=53.7%), was dominant in Russia, followed by p.Arg261Gln (AF=5.6%) and c.842C>T (p.Pro281Leu; AF=4.1%). The most common genotype was p.[Arg408Trp];[Arg408Trp] (30.6% of 458 different genotypes). In Southeastern Europe p.Arg408Trp was also the most frequent allele, but c.[143T>C];[1222C>T] / p.[Leu48Ser];[Arg408Trp] was the most prevalent genotype (GF=10.6%)

The distribution of patients with p.Arg408Trp on at least one allele (compound heterozygotes and homozygotes) decreased from 98% in Estonia to 89% in Poland, 76% in Russia, 69% in

Slovakia, 65% in Czechia, 40% in Austria, 36% in Germany, 10% in France, 6% in Italy, to only 4% in Spain (Table 3, Figure 5). Genotypes with the c.1066-11G>A splice site variant, occurred commonly in Armenia (48%), Turkey (32%), Iran (26%), Israel (21%), Spain (20%) and Italy (19%) (Table 3, Figure 5). The prevalence of p.Arg261Gln was 10–30% in most countries.

PKU phenotypes were more evenly distributed in Southern Europe (39.5% cPKU, 36.6% MHP, 23.9% mPKU), and milder phenotypes were more predominant, e.g. 37% had MHP in Italy. This was consistent with a high frequency of variants with a substantial residual PAH activity, e.g. c.1208C>T (p.Ala403Val; AF=8.4), c.898G>T (p.Ala300Ser; AF=3.6), c.734T>C (p.Val245Ala; AF=2.4) or c.1241A>G (p.Tyr414Cys; AF=2.4). Variants c.1066-11G>A and p.Arg261Gln, often called the “Mediterranean mutation” accounted for majority of mutant alleles, and the most frequent genotype was c.[1066-11G>A];[1066-11G>A] (GF=5.2%). In Portugal and Italy c.1066-11G>A and p.Arg261Gln occurred at similar rates (AF=16.2% and AF=10.7%, respectively). Variant c.1162G>A (p.Val388Met) was also common in Portugal (AF=11.3%) and Spain (AF=6.8%).

Low rates of cPKU were found in Southwestern Europe (42.7%) and Northern Europe (50.5%). In Southeastern Europe, Croatia (68%) and Bulgaria (65%) had a higher prevalence of cPKU than Serbia (48%). p.Arg408Trp was the most frequent variant in Croatia (AF=31.4%) and Bulgaria (AF=32.4%), but less common in Serbia (AF=14.7%). The p.Leu48Ser variant, which was initially identified in Turkey,<sup>21</sup> has the highest reported AF (31.3%) in Serbia. The most frequent genotypes in other countries locally were p.[Leu48Ser];[Leu48Ser] (14.7 %) in Serbia; p.[Leu48Ser];[Arg408Trp] (14.3 %) in Croatia; and p.[Arg408Trp];[Arg408Trp] genotype (16.2 %) in Bulgaria.

In Northern Europe c.1315+1G>A was the most common splice site variant (AF=25%). Most patients were compound heterozygous for p.[Arg408Trp];c.[1315+1G>A]. Denmark had the highest number of cases of the mild p.Tyr414Cys variant in the world and Sweden and Norway also a relatively high rate of mPKU (>43%). In Denmark the most frequent variant was c.1315+1G>A with an AF= 27.3%, while c.136G>A (p.Gly46Ser) occurred more commonly in Sweden and Norway. Variants p.Arg408Gln and c.896T>G (p.Phe299Cys) were specific for Norway.

The p.Arg408Trp variant was also the most frequent variant in all Central European countries, except for Switzerland where p.Arg261Gln (AF=15.3%) was more prevalent. Classical PKU was particularly frequent in Poland and Slovakia (>70%), but not in Czechia, consistent with a higher prevalence of the mild variant, p.Ala403Val. There was a wide spectrum of PKU

variants in Germany: of 102 distinct variants, p.Arg408Trp accounted for only 19.3%, and other variants such as p.Tyr414Cys (AF=9.7 %) as well as c.1315+1G>A (AF=8.8 %) were prominent. Almost half (44%) of patients in Germany had the mPKU phenotype, 32% had cPKU and 24% had MHP.

In France (1,307 patients) 59.1% had cPKU, and there were a total of 229 different variants, e.g. c.1066-11G>A (AF=7.4%), p.Arg261Gln (AF=6.5%), p.Arg408Trp (AF=5.5%), c.1315+1G>A (AF=4.5%) and c.838G>A (p.Glu280Lys; AF=3.7%). The p.Arg408Trp variant was also less common in the Netherlands (AF=2.2%). The predominant genotype in Western Europe was p.[Glu280Lys];[Glu280Lys].

### ***Latin America***

The most frequent variants here, p.Val388Met (AF= 13.9%), p.Arg261Gln (AF=10.7%) and c.1066-11G>A (AF=9.4%), were also prominent mutations in Southern Europe (see above). Homozygous p.[Val388Met];[Val388Met] (GF=4.4%) occurred most frequently. Of the three relevant South American countries, Brazil had the highest numbers of cPKU (63%), followed by Mexico (57%) and Argentina (43%). In comparison to Brazil or Argentina, Mexico's most frequent variant was c.60+5G>T (IVS1+5G>T).

### ***North America***

The most prevalent variants in North America also resembled those in European populations: p.Arg408Trp (18.5%), c.1066-11G>A (7.9%), c.1315+1G>A (6.9%), as did the genotype distribution p.[Arg408Trp];[Arg408Trp] (GF=4.0%)

### ***Middle East***

The predominance of c.1066-11G>A variant (AF=20.1%) and its homozygous genotype (GF=15.3%) was evident In Iran, Turkey, Israel and Saudi Arabia. Other frequent variants included p.Arg261Gln, c.168+5G>C (IVS2+5G>C), p.Pro281Leu and c.727C>T (p.Arg243\*). 81% of the Iranian patients had cPKU. In comparison to Iran, Turkey had a similar AF for c.1066-11G>A (22.9 %) and p.Arg261Gln (11.8 %), but strongly different phenotypes: 42% had cPKU, 35% MHP and 23% mPKU. This may be related to the augmented presence of mild variants such as p.Ala300Ser (7.4%) or c.1169A>G (p.Glu390Gly; 4.2%). In contrast to other Middle Eastern countries, the most frequent variant from Saudi Arabia was c.754C>T (p.Arg252Trp; 27,6%), followed by c.781C>T (p.Arg261\*; 11.2 %) and p.Val388Met (10.4 %); 30% of all patients had a p.Arg252Trp-associated genotype, i.e., compound heterozygous or homozygous (Table 3).

### ***Asia***

Most reported cases from Asia came from the east, with the most prevalent being the missense variant c.728G>A (p.Arg243Gln; AF= 21.8%), followed by c.611>G (9.8%) and c.721C>T (p.Arg241Cys; 8.7%). The most frequent genotype was homozygous p.[Arg243Gln],[Arg243Gln] (GF=5.6%).

In China, 62% of patients (n=2008) had a cPKU, 28% mPKU and 10% a MHP phenotype. A total of 234 different variants was reported, of which five, p.Arg243Gln, c.611>G, p.Arg241Cys, c.331C>T (p.Arg111\*) and c.1238G>C (p.Arg413Pro), had a frequency of >5% accounted for 52.5% of all alleles. Furthermore 679 different genotypes were reported, the most frequent being p.[Arg243Gln];[Arg243Gln]. Korea was the only country that, like China, had a high rate of cPKU (71%); this was less common in Japan (37%) and Taiwan (25%). The p.Arg243Gln, c.611>G, and p.Arg241Cys were detected in Korea, Taiwan, and China, each with allele frequencies >5%. The p.Arg111\* and p. Arg413Pro variants were common in Japan, China and Taiwan, but not Korea. The splice variant c.442-1G>A (IVS4-1G>A) was much more prevalent in Korea than in other Eastern Asian countries.

Genotypes with p.Arg243Gln were common in China (40%) and Korea (27%) (Table 3), while those with p.Arg241Cys were prevalent in Taiwan (41%), Japan (27%) and China (15%) (Table 3). Patients with the splice site variant c.442-1G>A were more common in Korea (18%), followed by China (8%) and Japan (7%). The splice variant c.611A>G (Ex6-96A>G), which masquerades as missense variant p.Tyr204Cys<sup>22</sup>, was found 19% of patients in China, 13 in Taiwan and 12 in Korea (Table 3).

### ***Africa***

The low number of reports available from Africa were considered not sufficiently representative of this world region. Thus, we did not analyze these data.

## **Discussion**

The aim of this study was to elucidate the prevalence of PKU and distribution of causative *PAH* variants worldwide and in different countries. We provided a rough estimation of global PKU prevalence by calculating the number of affected patients for countries based on the provided PKU prevalence and the total population in 2018<sup>23</sup> and an average prevalence for countries not employing newborn screening for PKU. Overall, it appears that there are about 0.45 million PKU patients worldwide, of whom at least two thirds have PKU that requires treatment (most patients had a severe, cPKU phenotype). An improvement in early diagnosis via NBS in countries still lacking it should be an urgent priority. With the multiple allelic combinations generated by so many mutations, a full description of all genotype/phenotype

correlations is impossible in this report. Nevertheless, it is clear that for many allelic combinations, often the most common in a region or country, such correlations are possible as detailed in the final three sections in the Results along with Table 2 and Figure S4.

Recognizing these relationships is important for counseling parents now that newborn screening predominates diagnosis of this disorder, particularly in instances when treatment recommendations are unclear (e.g., due to borderline blood Phe levels).

Although PKU has been found to be most common in European populations,<sup>9</sup> its similar prevalence in certain Middle Eastern countries (particularly Turkey and Iran) was a remarkable finding.<sup>24; 25</sup> A possible contributing factor may be the frequency of consanguinity in Islamic cultures, especially marriages between first-cousins, which would favor the autosomal recessive inheritance of PKU.<sup>26</sup> Prevalence and consanguinity was missing for many countries, making it more difficult to assess this hypothesis.<sup>27</sup>

We identified 758 different (previously reported) PKU variants in this study, emphasizing the strong genetic heterogeneity of PKU. Information of about 16,000 genotypes and phenotypes from PKU patients enabled a more detailed study of their distribution across different regions and countries than previously possible.<sup>28; 29</sup> Overall, the severe p.Arg408Trp variant was the most common, especially among Eastern European populations, in accordance with previous studies.<sup>30,31</sup> Furthermore, p.Arg408Trp dominates most Central European populations (Poland, Slovakia, Czechia, Austria, or Germany), again supporting previous studies.<sup>32-36</sup> Previous research had suggested that some PKU variants appeared to have been carried by migration, e.g. p.Arg408Trp, c.1066-11G>A, c.1315+1G>A or p.Phe299Cys and Arg408Gln.<sup>28; 37</sup> Our study supports and extends the knowledge on how certain variants appear to have spread within Europe and to different regions and countries worldwide. In particular, the distribution of PKU variants across Europe was consistent with successive waves of historical migration, and Figure 4 summarizes the likely geographical routes of transmission of some variants within Europe. It was previously suggested that the concomitant excess of (unaffected) PKU carriers is at least in part the result of over-dominant selection (“heterozygous advantage”).<sup>38</sup>

For example, Germany has occupied a ‘crossroads’ location during several migration waves throughout history, and displays a broad spectrum of *PAH* variants, with similarities to some other populations, such as northern European countries (Denmark, Sweden and Norway).<sup>36</sup> This implies a genetic connection of these regions, possibly during the Germanic settlement of Scandinavia. Norway’s most frequent variant, p.Gly46Ser, was not observed in the German



population. Furthermore, the occurrence of c.1066-11G>A in Germany was surprisingly high in our analysis, which might be explained by immigration from Turkey. This variant, often described as the “Mediterranean mutation”, may be of Italian origin<sup>39</sup> and has been found mainly in Southern European countries or the Middle-East.<sup>40,41</sup> Our study showed that c.1066-11G>A has the highest AF in Middle Eastern countries. A typical east-west gradient can be seen, originating from Western Asia and the Middle East, i.e., from Armenia (48%) via Turkey, Iran, Israel to Spain.

The p.Arg408Trp-associated genotype, the most common variant in people with Slavic roots, followed the east-central-southwest European axis, starting from Estonia and Russia, with the highest number of severe PKU patients, via Poland, Czechia, Slovakia, Germany, France, and Italy, down to Spain. The high prevalence of cPKU in Russia, and the dominance of the severe variant, p.Arg408Trp, were striking, given that Russia has a diverse ethnic composition, with as many as 160 ethnic groups.<sup>42</sup>

Another well-documented axis exists for the p.Tyr414Cys-associated genotype, from Northern to Western Europe, i.e. from Sweden, Norway and Denmark via Germany to the Netherlands. Some variants, e.g., c.1315+1G>A, seem to have spread within Scandinavian countries at high frequency. The same was true for the p.Glu390Gly-associated genotype concentrating between Slovenia, Croatia, Serbia and Austria. A trend were evident for very mild p.Ala300Ser-associated genotypes, frequent in Turkey and Israel, Italy to Spain. No trends for the distribution of p.Val245Ala and p.Ala403Val across Europe could be seen. The influence of migration patterns during history can also be seen in the heterogenous spectrum of *PAH* variants in Latin America. These countries were strongly marked by immigration of Europeans, especially from Southern Europe, during colonial times.<sup>43</sup> An example is p.Val388Met, which was especially prevalent in Brazil and Chile, once colonies of Portugal and Spain. In Argentina variants frequently occurring in Eastern, Central and Northern Europe; e.g., as p.Arg408Trp, p.Arg261Gln and c.1066-11G>A were more prevalent. In contrast to Brazil or Argentina, the Mexican population has a high proportion of indigenous individuals, which could be a reason for the exceptional high AF of the c.60+5G>T splice variant there.<sup>44</sup>

The USA has one of the most multi-racial and ethnic populations in the world. The strongest ancestral influence is European, mainly from Germany, Ireland, England, Italy and France<sup>45</sup>, which explains a comparable distribution of PKU variants in the USA and Europe.<sup>46</sup> African Americans have a much lower PKU incidence than Caucasian Americans.<sup>47</sup> It would be of

interest to follow up this study with regard to the different rates of growth of major ethnic groups (Caucasian, African-American, Hispanic, Asian).

While the overall PKU prevalence in China was 1:15,924, its distribution across the country varies significantly with higher rates in the north in comparison to the south.<sup>48</sup> In accordance with previous reports, p.R234Q accounts for 23% of all variants in China.<sup>49</sup> Three additional variants p.Arg243Gln, c.611A>G and p.Arg241Cys (AF>5%) were also common in Korea and Taiwan, indicating commonalities of migration movements and evolution among those nationalities. Prevalent PKU variants for Europe, Middle East, Latin America and the USA were uncommon in the Asian populations.

Despite immense progress in the diagnosis and treatment of PKU in the last decade, there are still too many areas of the world without adequate access to this care starting newborn screening. In particular, data from Africa, certain Asian and Southern American as well as Caribbean countries were missing.

Preliminary data indicated the great value of large patient databases with genotype and phenotype information and the introduction of the APV and GPV for the genotypic phenotype prediction.<sup>15</sup> This study extends these previous observations with additional information, and has confirmed the power of genotyping in prediction of phenotype and BH<sub>4</sub> responsiveness in PKU, thus offering a powerful tool of personalized medicine for this inherited metabolic disease.

This study also confirms that functionally mild variants (in a compound heterozygous constellation), with a substantial residual PAH activity are always dominant over inactive severe variants (null alleles) and give rise to a milder phenotype and often potential BH<sub>4</sub> responsiveness. Compared with severe classic PKU variants, milder variants were much rarer in number but determined the milder metabolic phenotype. Interestingly, homozygous mild variants have a higher APV and thus a milder phenotype when occurring in a compound heterozygous state with a null allele<sup>15</sup>. As a rule, two inactive severe variants are never BH<sub>4</sub>-responsive, although anecdotal reports sometimes suggest otherwise.<sup>50</sup>

In conclusion, this study provides an overview of the current distribution of the most important *PAH* variants and patients genotypes in various world regions. This information, together with the APV value, helps to predict the metabolic phenotype and the possible treatment options for PKU patients.

## Description of Supplemental Data

Supplemental Data include 4 Figures and 6 Tables.

**Figure S1.** Flow-chart of the literature search.

**Figure S2.** Structure of the *PAH* gene and variants tabulated in the locus-specific database (<http://www.biopku.org/pah/home.asp>). Regulatory domain: residues 1-142; catalytic domain: residues 143-410; oligomerization domain: residues 411-452. Cofactor (BH<sub>4</sub>)-binding regions (CBR): CBR1: residues 246-266; CBR2: residues 280-283; CBR3: residues 322-326; CBR4: residues 377-379. The RefSeq accession number for the *PAH* is ENSG00000171759; GeneBank: NM\_000277.1.

**Figure S3.** Heat maps of the distribution (%) of classic PKU (cPKU), mild PKU (mPKU) and mild hyperphenylalaninemia (MHP) A) in countries with at least 35 reported cases and B) heat maps of overall genotypes severity, expressed as genotypic phenotype value (GPV), The lower the GPV, more severe the phenotype. This Figure presents same information as Figure 4, but additional genotypic phenotype values (GPV) for Europe and other world regions in a more granular form and is thus complementary to Figure 4.

**Figure S4.** Heat maps of the distribution (%) of classic PKU (cPKU), mild PKU (mPKU) and mild hyperphenylalaninemia (MHP) in A) different European regions and C) in 7 world regions. Heat maps of overall genotypes severity, expressed as genotypic phenotype value (GPV), in B) European regions and D) in 7 world regions. The lower the GPV, more severe the phenotype. F) Ten most common variants, their allele frequency (AF) and corresponding allelic phenotype values (APV) for the 3 PKU phenotype categories. The RefSeq accession number for the *PAH* is ENSG00000171759; GeneBank: NM\_000277.1.

**Table S1.** Prevalence of PKU (all phenotypes) in world regions and countries and estimated number of PKU patients.

**Table S2.** *PAH* variants (n=758) found in the study cohort of PKU patients.

**Table S3.** Genotypes (n=3,659) found in the study cohort of PKU patients.

**Table S4.** Distribution of BH<sub>4</sub> responsiveness in different phenotype groups and descriptive analysis of maximal pretreatment blood Phe concentrations and BH<sub>4</sub> responsiveness. Patients classified as “slow responders” are included in the non-responsive group.

**Table S5.** Definition of countries in the world regions.

**Table S6.** Summary of the database information for all countries (Excel file).

## **Declaration of interests**

Nenad Blau has received honoraria and/or consulting fees from BioMarin Pharmaceuticals, Censa, Nestle Pharmaceuticals, and Homology Medicines. Alberto Burlina has received advisory board honoraria, speaker fees and travel support from Biomarin Pharmaceuticals, Nutricia, Cambrooke, PIAM, APR, Sanofi Genzyme, and Takeda. Barbara K. Burton has received honoraria and/or consulting fees from BioMarin, Shire (a Takeda company), Sanofi Genzyme, Ultragenyx, Alexion, Horizon, Denali, JCR Pharma, Regencbio, Inventiva, Chiesi, Homology Medicines, Aeglea, Agios, and Moderna. Georg F. Hoffmann received lecture fees from Takeda. Jerry Vockley received research funding from Biomarin Pharmaceuticals, Homology Pharmaceuticals, American Gene Technologies, Nestle Pharmaceuticals, Rubius Pharmaceuticals, and Synlogic Pharmaceuticals.

All other authors have no conflicts of interests and nothing to disclose.

## **Acknowledgements**

This article is dedicated in honor of Dr. Charles Scriver, the incomparable physician scientist, whose contributions to PKU and all metabolic disease have taught us the real meaning of medical research.

We are indebted to Suha Daas, who works at the National Newborn screening Program, Ministry of Health, Tel Hashomer Israel for her help and analyses. Farrah Rajabi, Fran Rohr, Ann Wessel, Leslie Martell, Steven Dobrowolski, Per Guldberg, and the late Flemming Guttler were greatly responsible for the data from New England. This work was funded in part by the Fundación Isabel Gemio-Fundación La Caixa (LCF/PR/PR16/11110018), and the Regional Government of Madrid (CAM, B2017/BMD3721). The research conducted at the Murdoch Children's Research Institute was supported by the Victorian Government's Operational Infrastructure Support Program. Funded by NIH grants R01DK117916 and R01NR016991. The work of Georg F. Hoffmann on newborn screening has been generously supported by the Dietmar Hopp Foundation, St. Leon-Rot.

A medical writer (Dr Mike Gwilt, GT Communications, UK) edited the authors' manuscript draft for conciseness, at the request of the Corresponding Author and funded by the Corresponding Author's institution.

## **Data and Code Availability**

The *PAH* variants dataset generated during this study is deposited in public *PAH*vdb repository at <http://www.biopku.org/pah/home.asp>.

The *PAH* genotypes dataset generated during this study is deposited in public BIOPKU repository at <http://www.biopku.org/biopku/home.asp>.

Specific export data can be obtained on request at the corresponding author.

## **Web Resources**

BIOPKU: <http://www.biopku.org/biopku/home.asp>

ClinVar: <https://www.ncbi.nlm.nih.gov/clinvar/>

GeneBank: <https://www.ncbi.nlm.nih.gov/genbank/>

HGMD: <http://www.hgmd.cf.ac.uk/ac/index.php>

HGVS nomenclature recommendations: <http://www.hgvs.org/content/guidelines>

LOVD: <https://www.lovd.nl/>

Mutalyzer 2.0: <http://www.mutalyzer.nl/>

OMIM: <https://omim.org/>

*PAH*vdb: <http://www.biopku.org/pah/home.asp>

## References

1. Blau, N., Van Spronsen, F.J., and Levy, H.L. (2010). Phenylketonuria. *Lancet* 376, 1417-1427.
2. Flydal, M.I., and Martinez, A. (2013). Phenylalanine hydroxylase: function, structure, and regulation. *IUBMB life* 65, 341-349.
3. Himmelreich, N., Shen, N., Okun, J.G., Thiel, C., Hoffmann, G.F., and Blau, N. (2018). Relationship between genotype, phenylalanine hydroxylase expression and in vitro activity and metabolic phenotype in phenylketonuria. *Mol Genet Metab* 125, 86-95.
4. Güttler, F. (1980). Hyperphenylalaninemia: Diagnosis and classification of the various types of phenylalanine hydroxylase deficiency in childhood. *Acta Paediat Scand* 280, 1-80.
5. Camp, K.M., Parisi, M.A., Acosta, P.B., Berry, G.T., Bilder, D.A., Blau, N., Bodamer, O.A., Brosco, J.P., Brown, C.S., Burlina, A.B., et al. (2014). Phenylketonuria scientific review conference: State of the science and future research needs. *Mol Genet Metab* 112, 87-122.
6. van Spronsen, F.J., van Wegberg, A.M., Ahring, K., Belanger-Quintana, A., Blau, N., Bosch, A.M., Burlina, A., Campistol, J., Feillet, F., Gizewska, M., et al. (2017). Key European guidelines for the diagnosis and management of patients with phenylketonuria. *Lancet Diabetes Endocrinol* 5, 743-756.
7. Vockley, J., Andersson, H.C., Antshel, K.M., Braverman, N.E., Burton, B.K., Frazier, D.M., Mitchell, J., Smith, W.E., Thompson, B.H., and Berry, S.A. (2014). Phenylalanine hydroxylase deficiency: diagnosis and management guideline. *Genet Med* 16, 188-200.
8. Blau, N., and Longo, N. (2015). Alternative therapies to address the unmet medical needs of patients with phenylketonuria. *Expert Opin Pharmacother* 16, 1-10.
9. Scriver, C.R. (2007). The PAH gene, phenylketonuria, and a paradigm shift. *Hum Mutat* 28, 831-845.
10. Gundorova, P., Zinchenko, R.A., Kuznetsova, I.A., Bliznetz, E.A., Stepanova, A.A., and Polyakov, A.V. (2018). Molecular-genetic causes for the high frequency of phenylketonuria in the population from the North Caucasus. *PLoS One* 13, e0201489.
11. Guldberg, P., Henriksen, K.F., Sipila, I., Guttler, F., and de la Chapelle, A. (1995). Phenylketonuria in a low incidence population: molecular characterisation of mutations in Finland. *J Med Genet* 32, 976-978.
12. Okano, Y., Kudo, S., Nishi, Y., Sakaguchi, T., and Aso, K. (2011). Molecular characterization of phenylketonuria and tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency in Japan. *Journal of human genetics* 56, 306-312.
13. Güttler, F., Guldberg, P., Henriksen, K.F., Mikkelsen, I., Olsen, B., and Lou, H. (1993). Molecular basis for the phenotypical diversity of phenylketonuria and related hyperphenylalaninaemias. *J Inherit Metab Dis* 16, 602-604.

14. Blau, N., Shen, N., and Carducci, C. (2014). Molecular genetics and diagnosis of phenylketonuria: state of the art. *Expert review of molecular diagnostics* 14, 655-671.
15. Garbade, S.F., Shen, N., Himmelreich, N., Haas, D., Trefz, F.K., Hoffmann, G.F., Burgard, P., and Blau, N. (2019). Allelic phenotype values: a model for genotype-based phenotype prediction in phenylketonuria. *Genet Med* 21, 580-590.
16. Zschocke, J. (2008). Dominant versus recessive: molecular mechanisms in metabolic disease. *J Inherit Metab Dis* 31, 599-618.
17. Scriver, C.R., and Waters, P.J. (1999). Monogenic traits are not simple: lessons from phenylketonuria. *Trends Genet* 15, 267-272.
18. Shen, N., Heintz, C., Thiel, C., Okun, J.G., Hoffmann, G.F., and Blau, N. (2015). Co-expression of phenylalanine hydroxylase variants and effects of interallelic complementation on in vitro enzyme activity and genotype-phenotype correlation. *Mol Genet Metab* 117, 328–335.
19. Evers, R.A.F.v.W., Annemiek M J ; Anjema, K., Lubout, C.M.A., van Dam, E., van Vliet, D., Blau, N., van Spronsen, F.J. (2020). The first European guidelines on phenylketonuria: its usefulness and implications for BH4 responsiveness testing. *J Inherit Metab Dis* 43, 244-250.
20. Muntau, A.C., Adams, D.J., Belanger-Quintana, A., Bushueva, T.V., Cerone, R., Chien, Y.H., Chiesa, A., Coskun, T., de Las Heras, J., Feillet, F., et al. (2019). International best practice for the evaluation of responsiveness to sapropterin dihydrochloride in patients with phenylketonuria. *Mol Genet Metab* 127, 1-11.
21. Konecki, D.S., Schlotter, M., Trefz, F.K., and Lichter-Konecki, U. (1991). The identification of two mis-sense mutations at the PAH gene locus in a Turkish patient with phenylketonuria. *Hum Genet* 87, 389-393.
22. Ellingsen, S., Knappskog, P.M., and Eiken, H.G. (1997). Phenylketonuria splice mutation (EXON6nt-96A-->g) masquerading as missense mutation (Y204C). *Hum Mutat* 9, 88-90.
23. United Nations -Population division, D.o.E.a.S.A. (2019). World population prospects, File Total population (both sexes combined) by region, subregion and country. Available at <https://population.un.org/wpp/>.
24. Matalon, R., and Michals, K. (1991). Phenylketonuria: screening, treatment and maternal PKU. *Clin Biochem* 24, 337-342.
25. El-Metwally, A., Al-Ahaidib, L.Y., Sunqurah, A.A., Al-Surimi, K., Househ, M., Alshehri, A., Da'ar, O.B., Razzak, H.A., and AlOdaib, A.N. (2018). The prevalence of phenylketonuria in Arab countries, Turkey, and Iran: A systematic review. *BioMed Res Internat* doi.org/10.1155/2018/7697210.
26. Shoraka, H.R., Haghdoost, A.A., Baneshi, M.R., Bagherinezhad, Z., and Zolala, F. (2020). Global prevalence of classic phenylketonuria based on neonatal screening programs: Systematic review and meta-analysis. *Clin Exp Pediatr* 63, 34-43.

27. Saadallah, A.A., and Rashed, M.S. (2007). Newborn screening: experiences in the Middle East and North Africa. *J Inherit Metab Dis* 30, 482-489.
28. Zschocke, J. (2003). Phenylketonuria mutations in Europe. *Hum Mutat* 21, 345-356.
29. Danecka, M.K., Woidy, M., Zschocke, J., Feillet, F., Muntau, A.C., and Gersting, S.W. (2015). Mapping the functional landscape of frequent phenylalanine hydroxylase (PAH) genotypes promotes personalised medicine in phenylketonuria. *J Med Genet* 52, 175-185.
30. Gundorova, P., Stepanova, A.A., Kuznetsova, I.A., Kutsev, S.I., and Polyakov, A.V. (2019). Genotypes of 2579 patients with phenylketonuria reveal a high rate of BH4 non-responders in Russia. *PLoS One* 14, e0211048.
31. Lillevali, H., Reinson, K., Muru, K., Simenson, K., Murumets, U., Mols, T., and Ounap, K. (2017). Hyperphenylalaninaemias in Estonia: Genotype-Phenotype Correlation and Comparative Overview of the Patient Cohort Before and After Nation-Wide Neonatal Screening. *JIMD Rep*, 10.1007/8904\_2017\_1061.
32. Bik-Multanowski, M., Kaluzny, L., Mozrzymas, R., Oltarzewski, M., Starostecka, E., Lange, A., Didycz, B., Gizewska, M., Ulewicz-Filipowicz, J., Chrobot, A., et al. (2013). Molecular genetics of PKU in Poland and potential impact of mutations on BH4 responsiveness. *Acta Biochim Pol* 60, 613-616.
33. Kadasi, L., Polakova, H., Ferakova, E., Hudecova, S., Bohusova, T., Szomolayova, I., Strnova, J., Hruskovic, I., Moschonas, N.K., and Ferak, V. (1995). PKU in Slovakia: mutation screening and haplotype analysis. *Hum Genet* 95, 112-114.
34. Reblova, K., Hrubá, Z., Prochazkova, D., Pazdirkova, R., Pouchla, S., and Fajkusova, L. (2013). Hyperphenylalaninemia in the Czech Republic: Genotype-phenotype correlations and in silico analysis of novel missense mutations. *Clin Chim Acta* 419, 1-10.
35. Sterl, E., Paul, K., Paschke, E., Zschocke, J., Brunner-Krainz, M., Windisch, E., Konstantopoulou, V., Moslinger, D., Karall, D., Scholl-Burgi, S., et al. (2013). Prevalence of tetrahydrobiopterine (BH4)-responsive alleles among Austrian patients with PAH deficiency: comprehensive results from molecular analysis in 147 patients. *J Inherit Metab Dis* 36, 7-13.
36. Zschocke, J., and Hoffmann, G.F. (1999). Phenylketonuria mutations in Germany. *Hum Genet* 104, 390-398.
37. Zschocke, J., Mallory, J.P., Eiken, H.G., and Nevin, N.C. (1997). Phenylketonuria and the peoples of Northern Ireland. *Hum Genet* 100, 189-194.
38. Krawczak, M., and Zschocke, J. (2003). A role for overdominant selection in phenylketonuria? Evidence from molecular data. *Hum Mutat* 21, 394-397.
39. Eisensmith, R.C., Okano, Y., Dasovich, M., Wang, T., Guttler, F., Lou, H., Guldberg, P., Lichter-Konecki, U., Konecki, D.S., Svensson, E., et al. (1992). Multiple origins for phenylketonuria in Europe. *Am J Hum Genet* 51, 1355-1365.



40. Okano, Y., Wang, T., Eisensmith, R.C., Longhi, R., Riva, E., Giovannini, M., Cerone, R., Romano, C., and Woo, S.L. (1991). Phenylketonuria missense mutations in the Mediterranean. *Genomics* 9, 96-103.
41. Kleiman, S., Avigad, S., Vanagaite, L., Shmuelewitz, A., David, M., Eisensmith, R.C., Brand, N., Schwartz, G., Rey, F., and Munnich, A. (1994). Origins of hyperphenylalaninemia in Israel. *Eur J Hum Genet* 2, 24-34.
42. Statistics, F.O.o. "On the results of the All-Russia Population Census 2010". Available at <https://www.rusemb.org.uk/russianpopulation/>.
43. Desviat, L.R., Perez, B., De Lucca, M., Cornejo, V., Schmidt, B., and Ugarte, M. (1995). Evidence in Latin America of recurrence of V388M, a phenylketonuria mutation with high in vitro residual activity. *Am J Hum Genet* 57, 337-342.
44. Vela-Amieva, M., Abreu-Gonzalez, M., Gonzalez-Del Angel, A., Ibarra-Gonzalez, I., Fernandez-Lainez, C., Barrientos-Rios, R., Monroy-Santoyo, S., Guillen-Lopez, S., and Alcantara-Ortigoza, M.A. (2015). Phenylalanine hydroxylase deficiency in Mexico: genotype-phenotype correlations, BH4 responsiveness and evidence of a founder effect. *Clin Genet* 88, 62-67.
45. Ramirez, K.R.H.N.A.J.R.R. (2011). Overview of Race and Hispanic Origin: 2010. United States Census Bureau. Available at <https://www.census.gov/library/publications/2011/dec/c2010br-02.html>.
46. Rajabi, F., Rohr, F., Wessel, A., Martell, L., Dobrowolski, S.F., Guldberg, P., Guttler, F., and Levy, H.L. (2019). Phenylalanine hydroxylase genotype-phenotype associations in the United States: a single center study. *Mol Genet Metab* 128, 415-421.
47. Hofman, K.J., Steel, G., Kazazian, H.H., and Valle, D. (1991). Phenylketonuria in U.S. blacks: molecular analysis of the phenylalanine hydroxylase gene. *American journal of human genetics* 48, 791-798.
48. Xiang, L., Tao, J., Deng, K., Li, X., Li, Q., Yuan, X., Liang, J., Yu, E., Wang, M., Wang, H., et al. (2019). Phenylketonuria incidence in China between 2013 and 2017 based on data from the Chinese newborn screening information system: a descriptive study. *BMJ Open* 9, e031474.
49. Li, N., Jia, H., Liu, Z., Tao, J., Chen, S., Li, X., Deng, Y., Jin, X., Song, J., Zhang, L., et al. (2015). Molecular characterisation of phenylketonuria in a Chinese mainland population using next-generation sequencing. *Sci Rep* 5, 15769.
50. Hennermann, J.B., Vetter, B., Kulozik, A.E., and Mönch, E. (2002). Partial und total tetrahydrobiopterin-responsiveness in classical and mild phenylketonuria (PKU). *J Inherit Metab Dis* 25, 21.

## Legends to figures

**Figure 1.** Prevalence of PKU (all phenotypes) in 64 countries from 6 world regions. For exact prevalence numbers see Table S1.

**Figure 2. A:** Boxplot (median, 25<sup>th</sup>-75<sup>th</sup> percentile, 1.5) of the maximal pretreatment blood Phe concentrations for three metabolic phenotypes in 6,369 PKU patients. The circles in the cPKU bar represent ordinary high blood Phe concentrations, since cPKU doesn't have an upper Phe limit for its classification; **B:** Contour plot of two-dimensional densities of pretreatment blood Phe concentrations and corresponding genotypic phenotype values (GPV) for 6,115 PKU patients;

**Figure 3. A:** Boxplot (median, 25<sup>th</sup>-75<sup>th</sup> percentile, 1.5) of GPV in 2,246 BH<sub>4</sub> non-responder and 1,755 responder PKU patients (for 2,114 out of 6,115 patients GVP was not known); **B:** Boxplot (median, 25<sup>th</sup>-75<sup>th</sup> percentile, 1.5) of GPV (APV<sub>max</sub>) for 11,584 PKU patients with a known genotype, but not tested for BH<sub>4</sub> responsiveness. Horizontal grey bar: separation area between GPVs for predicted BH<sub>4</sub> responsiveness (3.8-10) and non-responsiveness (0-3.3).

**Figure 4.** The world map with the relative frequency (%) of PKU and the corresponding most common variants for classic PKU (cPKU), mild PKU (mPKU) and mild hyperphenylalaninemia (MHP). Exact frequencies and additional genotypic phenotype values (GPV) for Europe and other world regions are presented in a more granular form in Figure S4. AF: allele frequency; APV: allelic phenotype value (cPKU=0-2.6; mPKU=2.7-6.6; MHP=6.7-10). The RefSeq accession number for the *PAH* is ENSG00000171759; GeneBank: NM\_000277.1. <sup>a</sup>c.611A>G reported as Ex6-96A>G splice variant.

**Figure 5.** Patterns of common *PAH* variants possibly associated with migration in Europe.

**Table 1.** Information included in the database

<b>Information provided</b>	<b>n (%)</b>
All patients	16,974
Complete genotype	16,196
- Compound heterozygotes	11,810 (73)
- Homozygotes	4,386 (27)
Phenotype and genotype	16,092
- cPKU	9,923 (62)
- mPKU	3,521 (22)
- MHP	2,648 (16)
Genotype and country	15,357 (91)
Blood Phe levels	6,371 (38)
Blood Phe levels and phenotype	6,369 (38)
Blood Phe levels and genotype	6,115 (36)
Phenotype, genotype and BH <sub>4</sub> responsiveness	5,597
- Non-responder	3,191 (57)
- Responder	2,316 (43)
Number of different genotypes	3,659
- Compound heterozygous	3,446 (94)
- Homozygous	213 (6)

cPKU: classic PKU; mPKU: mild PKU; MHP: mild hyperphenylalaninemia; BH<sub>4</sub>: tetrahydrobiopterin

**Table 2.** Summary of genotype and phenotype information for 51 countries with at least 35 patients reported.

Country	cPKU (%)	mPKU (%)	MHP (%)	No. 1 Frequent genotype	GF (%)	No. 2 Frequent genotype	GF (%)	No. 3 Frequent genotype	GF (%)	Different alleles (n)	No. 1 Frequent allele	AF (%)	No. 2 Frequent allele	AF (%)	No. 3 Frequent allele	AF (%)
<b>Europe</b>																
Austria	59.7	30.2	10.1	p.[Arg408Trp];[Arg408Trp]	7.0	p.[Arg261Gln];[ Arg408Trp]	6.3	c.[1315+1G>A];[1315+1G>A]	5.6	61	p.Arg408Trp	23.4	c.1315+1G>A	11.2	p.Arg261Gln	6.6
Bulgaria	64.9	27.0	8.1	p.[Arg408Trp];[Arg408Trp]	16.2	p.[Leu48Ser];[Arg408Trp]	8.1	p.[Leu48Ser];c.[1066-11G>A]	5.4	20	p.Arg408Trp	32.4	p.Arg261Gln	17.6	p.Leu48Ser	9.5
Croatia	67.7	19.1	13.2	p.[Leu48Ser];[Arg408Trp]	14.3	p.[Glu390Gly];[Arg408Trp6]	8.6	p.[Arg158Gln];[Arg408Trp]	5.7	26	p.Arg408Trp	31.4	p.Leu48Ser	14.3	p.Glu390Gly	7.1
Czechia	69.9	7.4	22.7	p.[Arg408Trp];[Arg408Trp]	18.2	p.[Arg408Trp];[Ala403Val]	4.0	p.[Arg158Gln];[Arg408Trp]	3.8	92	p.Arg408Trp	41.7	p.Ala403Val	5.3	p.Arg158Gln	4.6
Denmark	51.3	28.3	20.4	p.[Tyr414Cys];c.[1315+1G>A]	8.7	c.[1315+1G>A];[1315+1G>A]	7.7	p.[Arg408Trp];c.[1315+1G>A]	7.7	83	c.1315+1G>A	27.3	p.Arg408Trp	17.6	p.Tyr414Cys	12.1
Estonia	93.5	2.2	4.3	p.[Arg408Trp];[Arg408Trp]	66.3	p.[Arg408Trp];[Arg261Gln]	5.4	p.[Arg408Trp];[Leu48Ser]	4.3	16	p.Arg408Trp*	82.6	p.Arg261Gln	2.7	c.1315+1G>A	2.2
France	59.1	22.0	18.9	p.[Glu280Lys];[Glu280Lys]	2.2	p.[G352Valfs*48];[G352Valfs*48]	1.7	p.[Arg261Gln];[Arg261Gln]	1.5	229	c.1066-11G>A	7.4	p.Arg261Gln	6.5	p.Arg408Trp	5.5
Germany	32.0	44.0	24.0	p.[Arg408Trp];[Arg408Trp]	4.1	p.[Arg408Trp];[Tyr414Cys]	2.7	p.[Arg261Gln];[Arg261Gln]	2.4	102	p.Arg408Trp	19.3	p.Tyr414Cys	9.7	c.1315+1G>A	8.8
Italy	39.1	24.3	36.6	p.[Arg261Gln];c.[1066-11G>A]	2.8	c.[1066-11G>A];[1066-11G>A]	2.2	p.[Arg261Gln];[Arg261Gln]	1.8	159	p.Arg261Gln	10.7	c.1066-11G>A	10.6	p.Ala403Val	8.4
Netherlands	42.0	37.2	20.8	p.[Tyr414Cys];c.[1315+1G>A]	4.5	p.[Arg261Gln];c.[1315+1G>A]	3.6	p.[Pro281Leu];c.[1315+1G>A]	2.7	65	c.1315+1G>A	13.7	p.Arg261Gln*	8.5	p.Pro281Leu	7.6
Norway	52.1	43.8	4.2	p.[Gly46Ser];c.[1315+1G>A]	12.2	p.[Gly46Ser];c.[842+1G>A]	6.1	p.[Gly46Ser];[Phe299Cys]	6.1	24	p.Gly46Ser	16.3	p.Tyr414Cys	13.3	c.1315+1G>A	12.2
Poland	74.0	10.9	15.1	p.[Arg408Trp];[Arg408Trp]	40.0	p.[Arg408Trp];c.[1066-11G>A]	5.6	p.[Arg408Trp];[Arg158Gln]	4.4	94	p.Arg408Trp	64.6	c.1066-11G>A	4.3	p.Arg158Gln	3.5
Portugal	52.9	34.3	12.7	p.[Arg261Gln];[Arg261Gln]	7.8	c.[1066-11G>A];[1066-11G>A]	5.9	p.[Arg261Gln];[Val388Met]	4.9	34	c.1066-11G>A	16.2	p.Arg261Gln	16.2	p.Val388Met	11.3
Russia	81.0	9.9	9.1	p.[Arg408Trp];[Arg408Trp]	31.6	p.[Arg261Gln];[Arg408Trp]	5.0	p.[Pro281Leu];[Arg408Trp]	3.7	148	p.Arg408Trp	53.7	p.Arg261Gln	5.6	p.Pro281Leu	4.1
Serbia	48.0	28.0	24.0	p.[Leu48Ser];[Leu48Ser]	14.7	p.[Leu48Ser];[Arg408Trp]	10.7	p.[Leu48Ser];[ Arg158Gln]	4.0	30	p.Leu48Ser	31.3	p.Arg408Trp	14.7	p.Ile306Val	7.3
Slovakia	72.6	5.9	21.5	p.[Arg408Trp];[Arg408Trp]	26.5	p.[Arg408Trp];c.[1315+1G>A]	4.6	p.[Arg408Trp];[Ala403Val]	4.6	51	p.Arg408Trp	47.9	p.Arg158Gln	5.9	c.1315+1G>A	5.0
Slovenia	34.1	22.7	43.2	p.[Arg158Gln];[Glu390Gly]	9.1	p.[Arg261Gln];[Glu390Gly]	9.1	p.[Pro281Leu];[Ala403Val]	6.8	27	p.Arg408Trp	12.5	p.Glu390Gly	11.4	p.Arg158Gln	11.4
Spain	41.6	28.8	29.6	c.[1066-11G>A];[1066-11G>A]	5.2	p.[Ile65Thr];c.[1066-11G>A]	2.0	p.[Arg261Gln];c.[1066-11G>A]	2.0	155	c.1066-11G>A	11.1	p.Val388Met	6.8	p.Ile65Thr	6.4
Sweden	41.8	43.6	14.6	p.[Arg408Trp];c.[1315+1G>A]	16.4	p.[Arg408Trp];[Arg408Trp]	10.9	p.[Tyr414Cys];c.[1315+1G>A]	9.1	12	p.Arg408Trp	27.3	p.Tyr414Cys	23.6	c.1315+1G>A	21.8
Switzerland	54.8	30.6	14.6	p.[Arg261Gln];[Arg261Gln]	9.7	p.[Ile95del];[Tyr414Cys]	4.8	p.[Leu48Ser];[Leu48Ser]	4.8	39	p.Arg261Gln	15.3	p.Arg408Trp	8.1	c.1066-11G>A	7.3
<b>Asia</b>																
China	62.1	27.6	10.3	p.[Arg243Gln];[Arg243Gln]	6.1	p.[Arg243Gln];c.[611A>G] <sup>a</sup>	4.6	p.[Arg241Cys];[Arg243Gln]	3.7	234	p.Arg243Gln	23.3	c.611A>G <sup>a</sup>	10.2	p.Arg241Cys	8.0
Japan	34.5	40.0	25.5	p.[Arg413Pro];[Arg241Cys]	9.1	p.[Arg413Pro];[Arg413Pro]	7.3	p.[Tyr414Cys];c.[1315+1G>A]	7.3	31	p.Arg413Pro	18.2	p.Arg241Cys	16.4	p.Arg111*	7.3
Korea	71.0	24.7	4.3	p.[Arg241Cys];[Ala259Thr]	4.3	p.[Arg241Cys];[Arg243Gln]	4.3	c.[611A>G];[611A>G] <sup>a</sup>	3.2	47	p.Arg243Gln	13.3	c.442-1G>A	10.1	p.Arg241Cys	8.5
Taiwan	25.4	47.9	26.8	p.[Arg241Cys];[Arg241Cys]	8.5	p.[Arg241Cys];c.[611A>G] <sup>a</sup>	7.0	p.[Ala434Asp];[Arg408Trp]	2.8	34	p.Arg241Cys	24.6	p.Arg408Gln	12.0	c.611A>G <sup>a</sup>	7.7
<b>North America</b>																
USA	58.6	27.2	14.2	p.[Arg408Trp];[Arg408Trp]	4.0	c.[1066-11G>A];[1066-11G>A]	3.1	p.[Arg408Gln];c.[1315+1G>A]	2.5	214	p.Arg408Trp	18.5	c.1066-11G>A	7.8	c.1315+1G>A	6.9
<b>Latin America</b>																
Argentina	45.4	26.8	27.8	p.[Arg408Trp];[Arg408Trp]	4.1	p.[Arg408Trp];[Ala403Val]	3.1	p.[Val388Met];[Tyr414Cys]	3.1	47	p.Arg408Trp	10.3	p.Arg261Gln	9.8	c.1066-11G>A	9.3
Brazil	63.0	34.2	2.8	p.[Val388Met];[Val388Met]	5.0	p.[Arg261Gln];[Arg261Gln]	3.7	c.[1066-11G>A];[1066-11G>A]	2.3	56	p.Val388Met	16.2	p.Arg261Gln	11.6	c.1066-11G>A	9.4
Mexico	56.5	30.4	13.1	c.[60+5G>T];[60+5G>T]	10.6	p.[Val388Met];c.[60+5G>T]	10.6	c.[60+5G>A];[441+5G>T]	4.3	33	c.60+5G>T	23.4	p.Val388Met	9.6	c.1066-11G>A	7.4
<b>Middle East</b>																
Iran	81.1	10.3	8.6	c.[1066-11G>A];[1066-11G>A]	18.4	p.[Arg261Gln];[Arg261Gln]	7.6	c.[168+5G>C];[168+5G>C]	6.8	106	c.1066-11G>A	22.1	p.Arg261Gln	9.7	c.168+5G>C	8.7
Israel	50.5	21.7	27.8	c.[1066-11G>A];[1066-11G>A]	9.4	p.[Leu48Ser];[Leu48Ser]	5.2	c.[168+1G>A];[168+1G>A]	3.8	55	c.1066-11G>A	15.1	p.Leu48Ser	10.8	p.Ala403Val	7.3
Saudi Arabia	85.0	9.0	6.0	p.[Arg252Trp];[Arg252Trp]	25.4	p.[Val388Met];[Val388Met]	10.4	p.[Arg261*];[Arg261*]	9.0	21	p.Arg252Trp	27.6	p.Arg261*	11.2	p.Val388Met	10.4
Turkey	42.1	22.7	35.2	c.[1066-11G>A];[1066-11G>A]	13.9	p.[Arg261Gln];[Arg261Gln]	8.3	p.[Pro281Leu];[Pro281Leu]	4.6	64	c.1066-11G>A	22.9	p.Arg261Gln	11.8	p.Leu48Ser	8.8
<b>Australia</b>																
Australia	80.4	14.1	5.4	p.[Arg261Gln];c.[1315+1G>A]	7.4	p.[Arg408Trp];c.[1315+1G>A]	7.4	c.[1315+1G>A];[1315+1G>A]	6.3	37	c.1315+1G>A	23.2	p.Arg408Trp	20.5	p.Ile65Thr	10.5

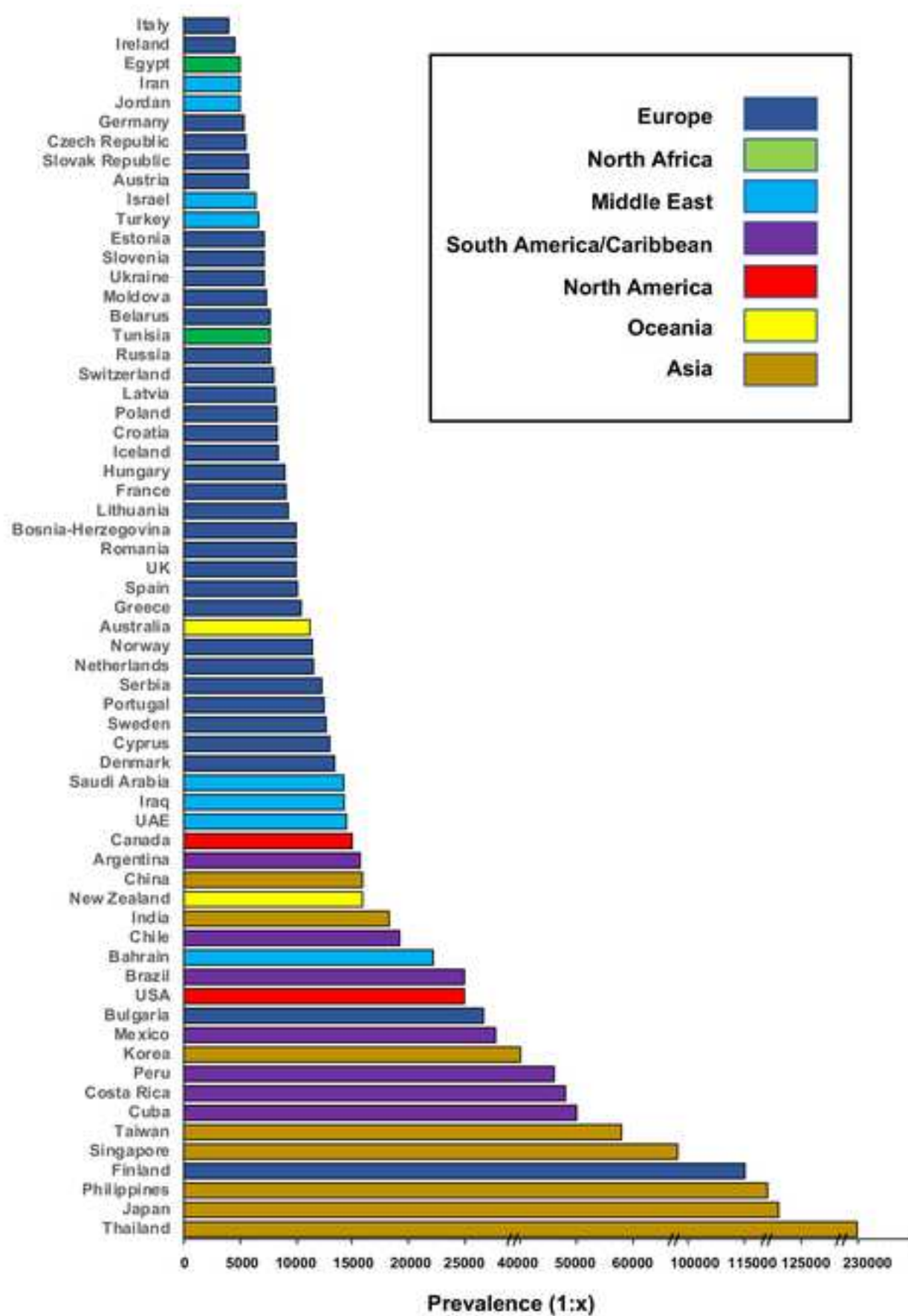
<sup>a</sup>c.611A>G reported as Ex6-96A>G splice variant. The RefSeq accession number for the *PAH* is ENSG00000171759; GeneBank: NM\_000277.1.

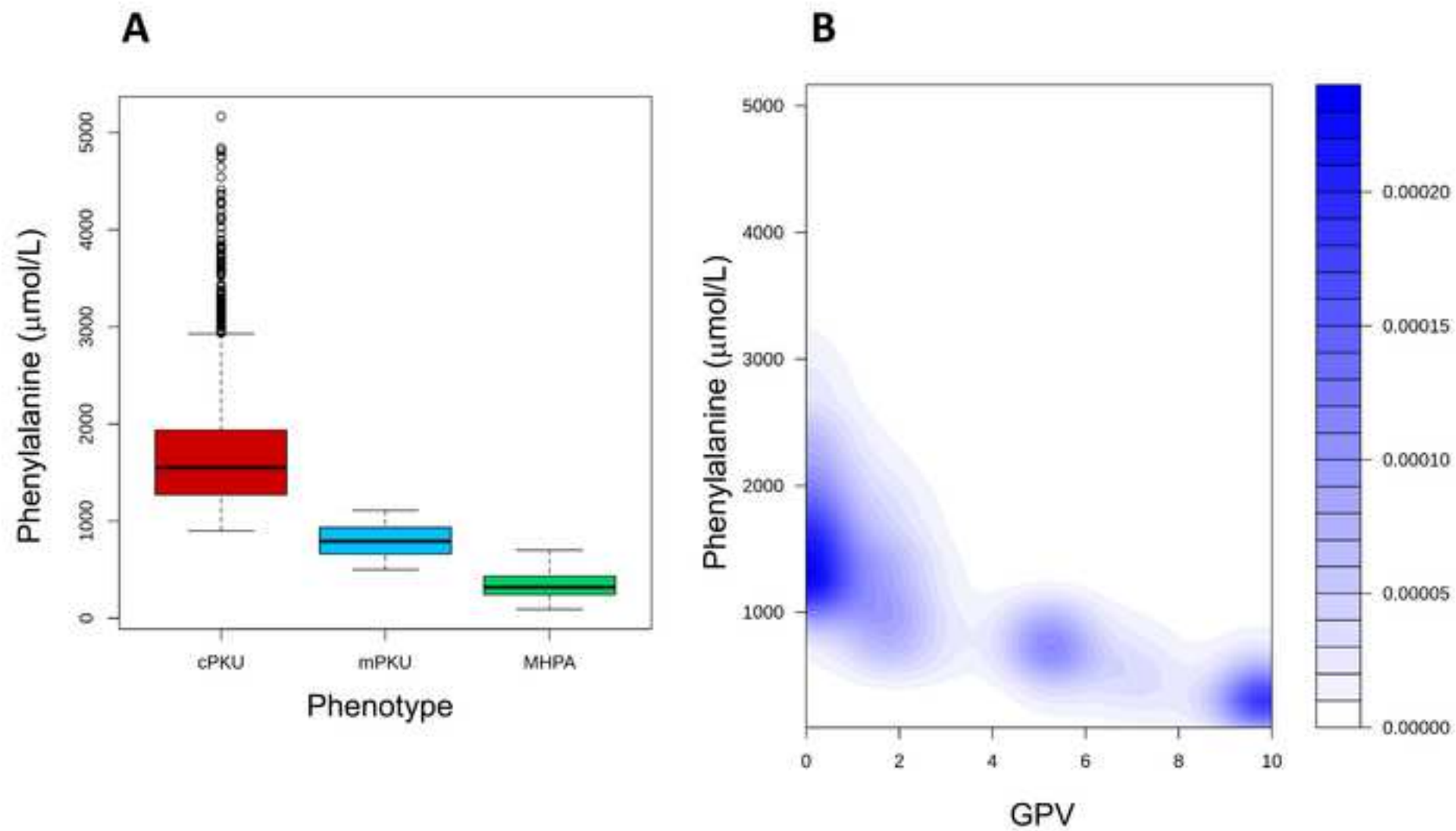
**Table 3.** Frequency (%) of patients with common variants on at least one allele (compound heterozygotes or homozygotes) in different countries. Source BIOPKU database.

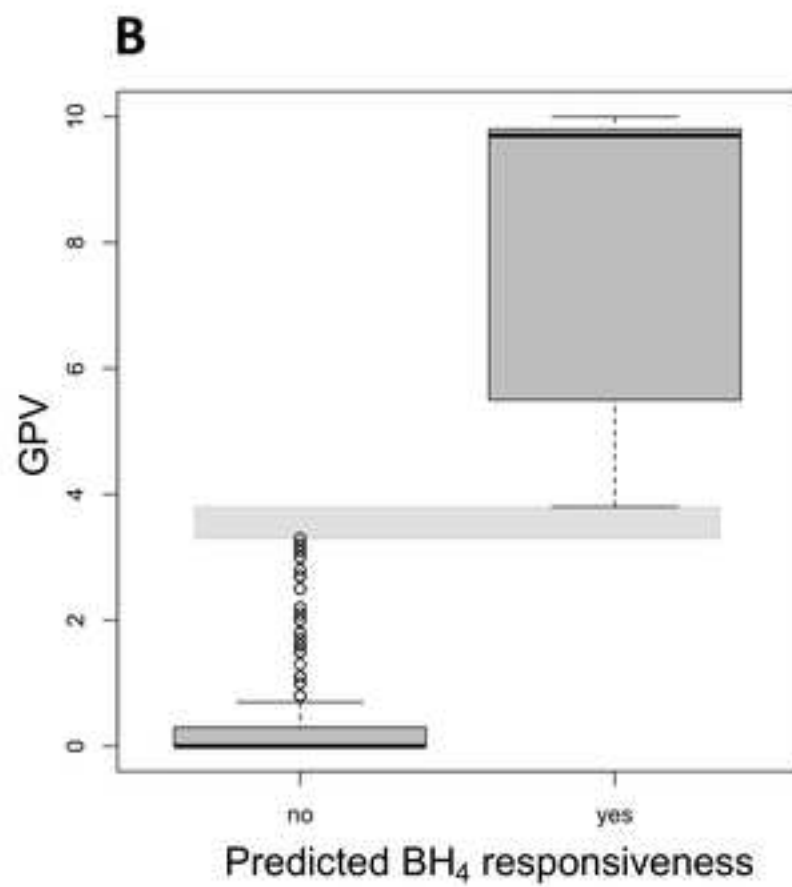
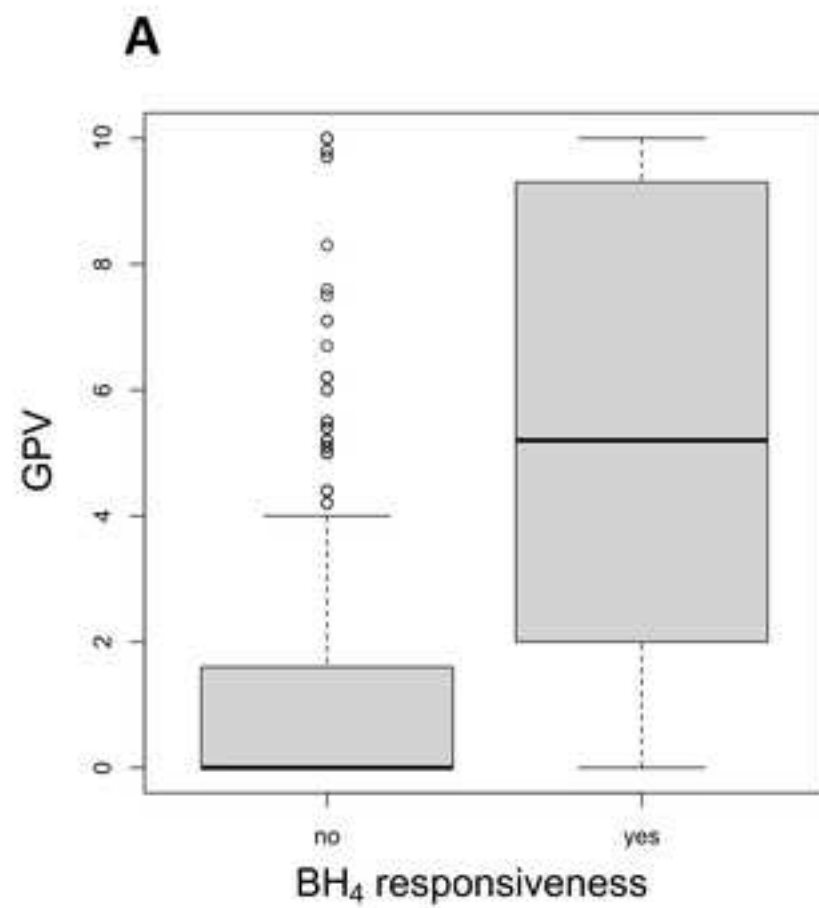
<b>p.Gly46Ser</b>	<b>%**</b>	<b>p.Leu48Ser</b>	<b>%</b>	<b>p.Ile65Thr</b>	<b>%**</b>	<b>p.Arg158Gln</b>	<b>%**</b>	<b>p.Arg241Cys</b>	<b>%**</b>	<b>p.Arg408Trp</b>	<b>%**</b>	<b>p.Arg261Gln</b>	<b>%**</b>
Norway	32.7	Serbia	48.0	Ireland	45.5	Slovenia	22.7	Taiwan	40.8	Estonia	98.9	Bulgaria	29.7
Denmark	2.0	Croatia	25.7	Australia	17.9	Italy	10.3	Japan	27.3	Poland	89.2	Portugal	23.5
Spain	1.8	Bulgaria	18.9	Brazil	12.8	Slovakia	10.0	China	15.0	Romania	87.5	Switzerland	21.0
Brazil	0.9	Israel	16.5	Spain	12.1	Netherlands	9.9	S. Korea	14.9	Russia	75.7	Brazil	20.5
USA	0.4	Turkey	13.4	Portugal	10.8	Czechia	8.6	Netherlands	4.5	Slovakia	69.4	Italy	19.7
										Czechia	64.6	Argentina	17.5
<b>p.Arg243Gln</b>	<b>%**</b>	<b>p.Val245Ala</b>	<b>%</b>	<b>p.Arg252Trp</b>	<b>%**</b>	<b>p.Glu280Lys</b>	<b>%**</b>	<b>p.Pro281Leu</b>	<b>%**</b>	Croatia	57.1	Netherlands	15.7
China	40.4	Italy	4.9	Saudi Arabia	29.9	France	5.1	Netherlands	14.8	Bulgaria	48.6	Turkey	15.3
S. Korea	26.6	Slovakia	2.3	Brazil	9.6	USA	4.3	Turkey	11.2	Sweden	43.6	Slovenia	13.6
USA	19.7	Germany	2.1	Portugal	6.9	Australia	4.2	Portugal	10.8	Austria	39.9	Germany	13.6
Taiwan	11.3	Denmark	1.5	Slovakia	6.8	Spain	3.5	Italy	10.3	Germany	35.7	Croatia	12.9
Spain	5.1	France	1.2	Czechia	5.1	Russia	2.5	Germany	9.4	Australia	34.7	Austria	12.6
										USA	32.9	Spain	11.9
<b>p.Phe299Cys</b>	<b>%**</b>	<b>p.Ala300Ser</b>	<b>%</b>	<b>p.Val388Met</b>	<b>%**</b>	<b>p.Glu390Gly</b>	<b>%**</b>	<b>p.Ala403Val</b>	<b>%**</b>	Denmark	28.8	Iran	11.8
Norway	14.3	Turkey	13.0	Chile	52.4	Slovenia	22.7	Italy	15.6	Serbia	28.0	Australia	11.6
Australia	4.2	Israel	9.0	Brazil	28.8	Croatia	14.3	Slovenia	13.6	Slovenia	22.7	France	11.5
USA	2.3	Italy	7.1	Portugal	20.6	Serbia	10.7	Israel	13.2	Argentina	16.5	Russia	10.3
France	1.1	Iran	3.8	Mexico	19.1	Turkey	6.5	Argentina	10.3	Switzerland	14.5	Serbia	9.3
Russia	0.2	Spain	3.7	Spain	13.1	Austria	5.6	Czechia	10.1	France	9.9	Israel	9.0
										Turkey	8.8	USA	8.6
<b>p.Tyr414Cys</b>	<b>%**</b>	<b>c.611A&gt;G*</b>	<b>%</b>	<b>c.442-1G&gt;A</b>	<b>%**</b>	<b>c.1066-11G&gt;A</b>	<b>%**</b>	<b>c.1315+1G&gt;A</b>	<b>%**</b>	Israel	7.1	Slovakia	6.8
Sweden	38.2	China	18.7	Korea	18.1	Armenia	47.8	Denmark	46.9	Brazil	6.8	Estonia	5.4
Norway	24.5	Taiwan	12.7	China	8.0	Turkey	31.9	Sweden	36.4	Italy	5.7	Poland	3.6
Denmark	22.2	S. Korea	11.7	Japan	7.3	Iran	25.7	Norway	26.0	Netherlands	4.5	Czechia	2.3
Germany	18.0	France	0.2	Poland	0.4	Israel	20.8	Netherlands	22.4	Spain	4.0	Denmark	2.3
Netherlands	11.7	Russia	0.1	Russia	0.1	Spain	20.1	Germany	17.4	Iran	1.4	China	1.6

\*c.611G>A reported as Ex6-96A>G splice variant; \*\*Percentage of patients (of a total number of patients) with a variant occurring in a compound heterozygous or homozygous constellation. For a total number of patients in each country see Table S6. The RefSeq accession number for the *PAH* is ENSG00000171759; GeneBank: NM\_000277.1.

Figure 1

[Click here to access/download;Figure;Figure 1 AJHGR2.tif](#)







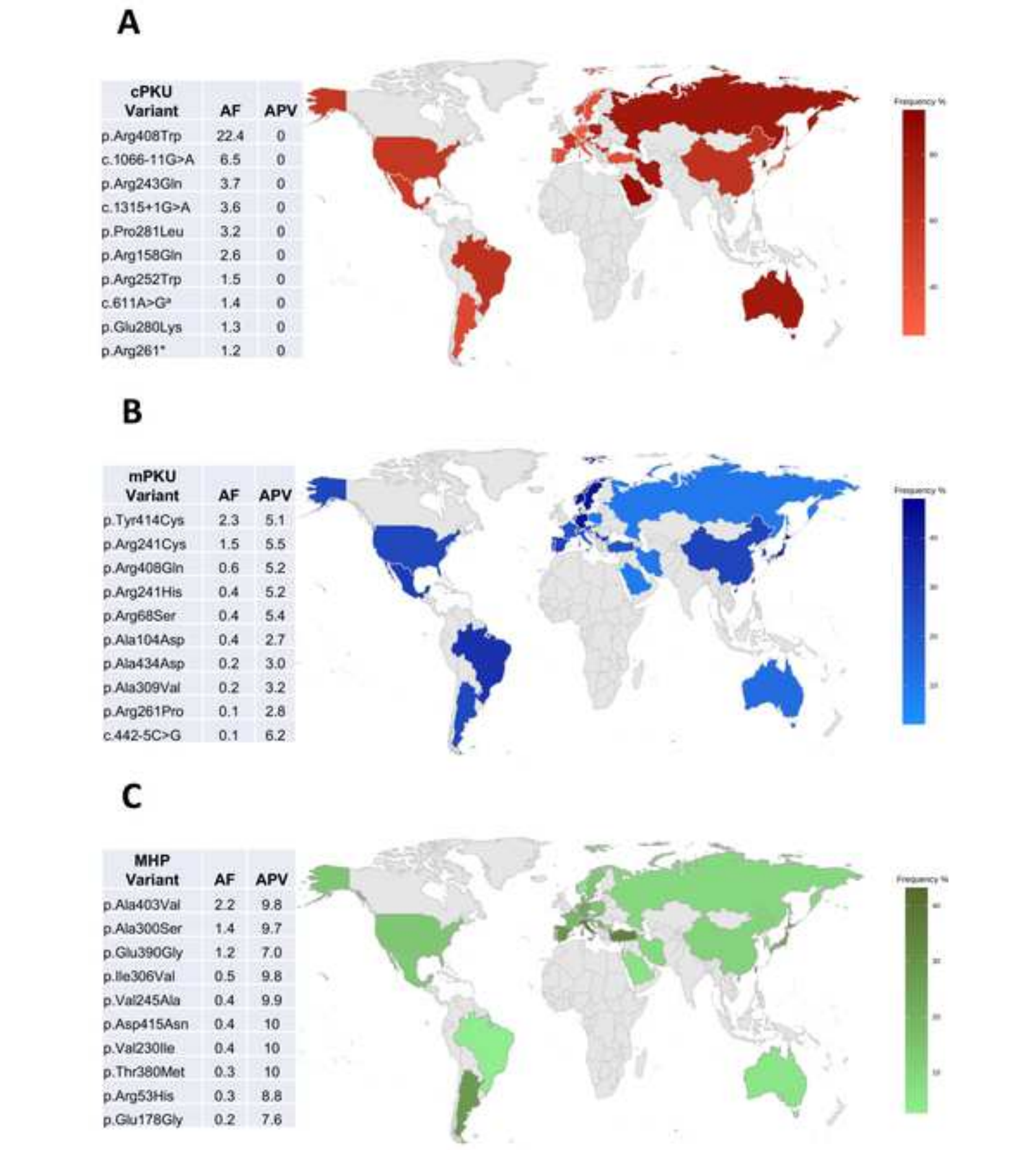


Figure 5

[Click here to access/download;Figure;Figure 5 AJHGR2.tif](#)

